Smart temperature sensors in standard CMOS

K.A.A. Makinwa*

Electronic Instrumentation Laboratory / DIMES, Delft University of Technology, Delft, The Netherlands

Abstract

A smart temperature sensor is an integrated system consisting of a temperature sensor, its bias circuitry and an analog-to-digital converter (ADC). When manufactured in CMOS technology, such sensors have found widespread use due to their low cost, small size and ease of use. In this paper the basic operating principles of CMOS smart temperature sensors are explained and the state-of-the-art is reviewed. Two new figures of merit for smart temperature sensors are defined, which express the tradeoff between their energy/conversion and their resolution and inaccuracy, respectively. A survey of data published over the last 25 years shows that both these figures of merit usefully bound the performance of state-of-the-art smart temperature sensors.

© 2010 Published by Elsevier Ltd.

Keywords: Smart temperature sensor; temperature sensor; survey; figure of merit;

1. Introduction

Temperature is one of the most often-measured environmental quantities. This is because most physical, electronic, chemical, mechanical and biological systems exhibit some form of temperature dependence. As a consequence, temperature measurement and control are critical tasks in many applications.

As shown in Fig. 1, a smart temperature sensor is an integrated system that consists of a temperature sensor and its interface electronics, i.e. bias circuitry and an analog-to-digital converter (ADC). Smart temperature sensors manufactured in standard CMOS technology have many advantages, such as low-cost, small size and ease of use [1]. Since most of the physical properties of silicon are temperature dependent, there is no shortage of CMOS-compatible devices, e.g. transistors and resistors, which could, potentially, be used as temperature sensors. In this role, however, most of these devices are far from ideal. They typically output small analog signals (millivolt range), which, moreover, are sensitive to process spread and (packaging) stress. These signals must then be digitized, preferably with respect to an accurate on-chip reference, by precision interface electronics. In consequence, the design of accurate CMOS smart temperature sensors is quite challenging.

* Corresponding author. Tel.: +31-15-2786466; Fax: +31-15-2785755.
E-mail address: k.a.a.makinwa@tudelft.nl

1877-7058 © 2010 Published by Elsevier Ltd.
doi:10.1016/j.proeng.2010.09.262
Fig. 1. Block diagram of an integrated smart temperature sensor

2. CMOS-compatible temperature sensors

The temperature dependence of transistors means that they can be used as temperature sensors. In CMOS technology, the same diffusions normally used to realize MOSFETs can be used to realize parasitic BJTs. While smart temperature sensors based on lateral PNP transistors have been realized [2][3], nowadays the vertical PNP transistor (Fig. 2) is usually preferred because of its relative insensitivity to process spread and packaging stress [4][5][6]. In modern processes with twin well or deep n-well options, vertical NPN transistors can also be made [7].

Fig. 2. Cross-section of a vertical PNP transistor realized in a standard CMOS process

The temperature-sensing accuracy of BJTs is limited by the effect of process spread on their saturation current ($I_s$), which can lead to errors of a few degrees. Fortunately, such errors can be reduced to the ±0.1°C (3σ) level by a single room-temperature calibration [8]. In the case of MOSFETs, both the threshold voltage ($V_T$) and the transconductance coefficient ($g_m$) are affected by spread. As a result, they must usually be calibrated at two temperatures in order to approach the accuracy of BJTs. Since the extra cost associated with such two-point calibration can be quite significant, most commercial CMOS temperature sensors are based on BJTs.

On-chip resistors can also be used as temperature sensors (thermistors). Since lightly doped n-well and polysilicon resistors can have large temperature coefficients (up to a few percent per °C), they can be used to realize compact and energy efficient smart sensors [9][10][11]. Compared to transistors, however, the temperature
dependence of such thermistors is typically not well defined, which means that they must be calibrated at two or even three temperatures to achieve similar accuracy.

Temperature can also be sensed by measuring the thermal diffusivity, i.e. the rate of heat diffusion, of silicon [12][13][14]. Since IC-grade substrate silicon is highly pure (lightly doped), its thermal diffusivity is quite insensitive to process spread.

As shown in Fig. 3, a thermal diffusivity sensor can be made by realizing a heater and a (relative) temperature sensor in the same silicon substrate. Electrical signals applied to the heater induce local temperature variations in the substrate, through which they diffuse before being detected by the temperature sensor. For this reason, the structure is also known as an electro-thermal filter (ETF) [13]. The process of diffusion is associated with a thermal delay, which is determined by the ETF’s geometry (the distance \( r \) in Fig. 1), and by the thermal diffusivity of the silicon substrate. Since lithographic tolerances will cause variations in the geometry, ETFs also suffer from process spread. However, when used as temperature sensors, they have been shown to exhibit inaccuracies of less than 0.2°C after a low-cost batch calibration [15].

3. Calibration and trimming

The accuracy and linearity of all CMOS temperature sensors can be improved by calibrating and then trimming them at one or more known temperatures. Since the thermal time constant of a packaged sensor is typically in the order of a few seconds, the thermal calibration of individual devices can be a time consuming and, thus, expensive procedure. Calibration at wafer level is significantly cheaper, but is less accurate, since the stress associated with (plastic) packaging will introduce errors in the order of a few tenths of a degree. Although stress-relieving coatings can be used, these also increase production costs. Another low-cost alternative is batch calibration, in which parameters determined from the calibration of a few samples are used to trim an entire batch of devices. Batch calibration of a BJT-based sensor resulted in an inaccuracy of ±0.25°C (3\( \sigma \)) over the military range (–55°C to 125°C), which improved to ±0.1°C (3\( \sigma \)) after individual calibration [16].

It should be noted that the effect of process spread on a temperature sensor’s accuracy is random, and so measurements on many samples are required to obtain reliable estimates of its inaccuracy. For a Gaussian distribution the standard deviation estimated from measurements on \( N \) samples will itself have a standard deviation of \( 1/\sqrt{2N} \), about 18% in the case where \( N=16 \).

4. Interface electronics

Assuming a target resolution of 0.01°C resolution and a linear sensor characteristic, an ADC with at least 15-bit resolution is required to cover the military temperature range (–55°C to 125°C). Furthermore, the ADC must achieve 11-bit accuracy in order to keep its own errors below the 0.1°C level. Since packaged chips typically have thermal time constants of a few seconds, ADCs for temperature sensors can be quite slow, with typical conversion rates of less than 100Hz. Since sigma-delta (\( \Sigma \Delta \)) ADCs can trade-off low speed for high resolution, they are widely used in smart temperature sensors. The required accuracy can be achieved through the use of dynamic techniques.
such as auto-zeroing, correlated-double sampling, chopping and dynamic element matching (DEM) [17][18]. In applications where only moderate resolution is required, successive-approximation ADCs have also been used [19].

Fig. 4. Block diagram of a bandgap temperature sensor

5. BJT-based temperature sensors

The block diagram of a typical BJT-based temperature sensor is shown in Fig. 4. The heart of the sensor consists of two vertical PNPs biased at different collector current densities. Over a few decades of collector current $I_C$, the base-emitter voltage $V_{BE}$ of a BJT may be expressed as:

$$V_{BE} = kT/q \ln(I_C/I_S)$$

(1)

Since the saturation current $I_S$ has a strong positive temperature dependence, $V_{BE}$ has a negative temperature dependence, which is almost linear and has a slope of approximately $-2\text{mV/}^\circ\text{C}$. However, $I_S$ is quite process dependent, and as a result $V_{BE}$ is sensitive to process spread. Fortunately, the effect of spread can be effectively compensated for by trimming $I_C$ and thus normalizing the $I_C/I_S$ term in (1). The PNPs are biased via their emitters, and so, depending on their current gain, extra base-current compensation circuitry may be needed to achieve well-defined collector currents [8].

In contrast, the difference in the base-emitter voltages of the two transistors $\Delta V_{BE}$, is process independent:

$$\Delta V_{BE} = (kT/q)\ln(p)$$

(2)
For a typical collector current-density ratio $p = 5$, $\Delta V_{BE}$ will have a positive temperature coefficient of $\approx 140\mu V/°C$.

As shown in Figs. 4 and 5, both $V_{BE}$ and $\Delta V_{BE}$ can be combined with the help of an appropriate gain factor $\alpha$ (typically about 16 for $p=5$) to generate a temperature-dependent voltage $V_{PTAT} = \alpha \Delta V_{BE}$ and a temperature-independent reference voltage $V_{REF}$. These voltages can then be applied to a sigma-delta ADC, whose output code $\mu$ will then be a digital representation of temperature. $V_{REF}$ is roughly equal to the bandgap voltage of silicon (about 1.2V), and as a result, sensors based on the Fig. 4 architecture are often referred to as bandgap temperature sensors.

Bandgap temperature sensors can cover the temperature range from -70°C to 150°C [3][16][20]. Depending on the process used, their untrimmed accuracy is then in the order of a few degrees [21][22]. Over the military temperature range and when packaged in low-stress ceramic packages, they can achieve inaccuracies of $\pm 0.1°C$ (3σ) after one-point calibration [8][16]. Plastic-packaged devices can achieve an inaccuracy of $\pm 0.15°C$ (3σ) over a somewhat smaller range: $-20°C$ to $105°C$ [23].

An alternative method of calibrating bandgap temperature sensors involves determining the die temperature by measuring $\Delta V_{BE}$ with the help of accurate external electronics. Such “electrical” calibration is nearly as accurate as thermal calibration, but does not require accurate thermal settling, and so can be carried out in much less time [24].

### 6. MOSFET-based temperature sensors

When operated in the sub-threshold region, the drain current and the gate-source voltage $V_{GS}$ of a MOSFET have an exponential relationship:

$$V_{GS} - V_{TH} = (\eta kT / q) \ln(I_D / I_0)$$  \hspace{1cm} (3)

where $I_0$ is a process-dependent parameter and $\eta$ is the sub-threshold slope factor. Its similarity to (1) means that the BJTs in the architecture of Fig. 2 can be replaced by MOSFETs [25]. After a one-point calibration, a MOSFET-based $V_{PTAT}$ generator achieved an inaccuracy of about $\pm 2°C$ from $10°C$ to $80°C$ [26].

Alternatively, time-to-digital converters (TDCs) have been proposed, in which the time it takes for the sub-threshold current to charge a capacitor to a fixed threshold is measured. After an initial batch calibration to derive a master curve, a low-cost one-point calibration of individual devices was sufficient to achieve inaccuracies of a few
degrees over from 40°C to 90°C [27]. By using two-point calibration, the inaccuracy can be improved to about ±1°C from -10°C to 30°C [28]. It should be noted that in contrast to band-gap temperature sensors, which are self-referenced, TDCs typically require an external timing reference, which often takes the form of a signal derived from an external quartz-crystal clock.

Temperature sensors based on the temperature-dependent propagation delay of a chain of inverters have also been proposed [29]. The basic architecture is shown in Fig. 6. Alternatively, the frequency of a ring oscillator can be measured. The average propagation delay $T_P$ of an inverter composed of PMOS and NMOS devices of equal strength, and which drives a load capacitance $C_L$ may be expressed as [29]:

$$T_P = \frac{(L/W) C_L}{\mu C_{ox} (V_{oo} - V_{in})} \ln \left( \frac{3V_{thy} - 4V_{th}}{V_{oo}} \right)$$  \hspace{1cm} (4)

In this equation, the temperature dependent parameters are the mobility $\mu$ and the threshold voltage $V_{th}$. If the supply voltage $V_{dd}$ is constant and much larger than $V_T$, then the temperature dependency of $T_P$ will be mainly determined by the mobility, which varies as

$$\mu = \mu_0 \left( \frac{T}{T_0} \right)^{km}$$  \hspace{1cm} (5)

where $T$ and $T_0$ are the actual and reference temperatures and $km = -1.2$ to $-2.0$. As a result, an inverter’s delay has a positive temperature coefficient. From (4), however, this delay is also sensitive to variations in $V_{dd}$ and $V_{th}$. At a fixed supply voltage and after a two-point calibration, inaccuracies of about ±0.5°C (3σ) have been achieved from 0°C to 90°C [30]. However, the corresponding supply sensitivity is about 10°C/V at room temperature, which is more than an order of magnitude worse than that of band-gap temperature sensors. Alternatively, an individual one-point calibration can be combined with a master curve obtained by batch calibration. The resulting inaccuracy, however, is then only about ±3°C from 0°C to 100°C [31].

Compared to BJT-based temperature sensors, the main advantage of MOSFET-based sensors is that their temperature dependence is very well modelled, even in baseline CMOS processes. However, their useful operating range appears to be limited to only about 100°C, which is about half that of BJT-based temperature sensors. Furthermore, MOSFET-based sensors typically require a more costly two-point calibration to achieve comparable accuracy.
7. Thermal diffusivity sensors

Unlike the transistor-based sensors described above, thermal diffusivity sensors are inherently time-domain devices. When driven at a constant frequency $f_{ref}$, the phase shift of the ETF shown in Fig. 3, can be expressed as:

$$\phi_{ETF} \propto r \frac{f_{ref}}{D}$$

(6)

where $D$ is the thermal diffusivity of silicon [32]. Over the military temperature range, $D$ can be approximated by a power law: $D \propto 1/T^n$, where $n \approx 1.8$ [33][34]. As a result,

$$\phi_{ETF} \propto \left(r \sqrt{f_{ref}} \right) T^{n/2}$$

(7)

Since $n/2 \approx 0.9$, the ETF’s phase shift will be a near-linear function of temperature. The residual non-linearity, being systematic, can easily be compensated for.

In an ETF, the distance between the heater and the relative temperature sensor should be maximized to minimize the effects of lithographic spread. Since silicon is a good conductor, and heater power dissipation is limited to a few milliwatts by the need to minimize self-heating errors, this means that the temperature fluctuations at the sensor will be quite small. In the case when the sensor is implemented as a thermopile, its output signals will typically be at the sub-millivolt level [13]. Fortunately, the phase of such small signals can be digitized with sufficient resolution by a phase-domain sigma-delta TDC [35][36]. Resolutions of 0.01°C have been achieved into a 1Hz bandwidth [15].

Compared to transistor-based temperature sensors, thermal diffusivity sensors are rather slow and dissipate significantly more power. However, being time based, they are insensitive to leakage current and so can operate up to 160°C [37]. They are also insensitive to process spread, as well as to the mechanical stress associated with plastic packaging [15]. As a result, high accuracy can be achieved with a low-cost wafer-level batch calibration.

8. Figures of merit

Given the variety of smart temperature sensors that can be realized in standard CMOS, it would be useful to devise a single figure of merit (FOM) to express their performance. Since a smart temperature sensor may be regarded as temperature-to-digital converter, an ADC FOM, involving energy per conversion and resolution could be defined [38]. It should be noted, however, that other figures of merit involving, for instance, chip area or cost of calibration, may be more suitable for specific applications.

Based on a survey of performance data published over the last 25 years [39], the energy per conversion versus the resolution of several smart temperature sensors is plotted in Fig. 7. It can be seen that the product of energy/conversion and the square of resolution is a FOM that usefully bounds the state-of-the-art. This is not too surprising, since temperature sensors typically output small signals and, in turn, the resolution of smart temperature solutions is often limited by thermal noise.

Depending on the intended application, the inaccuracy of temperature sensors is specified over many different temperature ranges. To normalize for this, the relative inaccuracy of a temperature sensor can be defined as its peak-to-peak inaccuracy divided by the temperature range. This is similar to the “box method” often used to express the temperature dependency of voltage references. With this definition a FOM can be defined as the product of energy/conversion and the square of relative inaccuracy. This is plotted in Fig. 8 along with the energy/conversion and relative inaccuracy of several smart temperature sensors [39]. It can be seen that the inaccuracy FOM does a good job of bounding the state-of-the-art. At first glance, this is rather surprising, since the inaccuracy of a
A temperature sensor over a given temperature range is mainly determined by its immunity to process spread and not by its energy consumption. However, in order to benefit from a given inaccuracy a commensurate resolution is required, and so the inaccuracy FOM is indirectly related to the resolution FOM.

![Fig. 7. Smart temperature sensor performance data: energy per conversion versus resolution](image1)

![Fig. 8. Smart temperature sensor performance data: energy per conversion versus relative inaccuracy](image2)
9. Outlook

Research into smart temperature sensors is currently driven by the requirements of a few key applications: the monitoring of food and other perishable with smart RFID tags, the temperature compensation of MEMS resonators in frequency references and the thermal management of large SoCs such as microprocessors. In the case of smart RFID tags, the emphasis is on low-cost and on ultra-low energy/power consumption. Micropower smart temperature sensors based on BJTs (microwatts) [40][41][42] and MOSFETs (nanowatts) [28][43][44] have been demonstrated. MEMS resonators are inherently temperature dependent and so MEMS frequency references employ temperature compensation schemes based on embedded smart temperature sensors. Since the temperature sensor’s inaccuracy and noise will limit the frequency reference’s inaccuracy and near-to-carrier phase noise, this application requires sensors with both high accuracy and high-resolution [9][45]. In the case of SoC thermal management, the emphasis is on achieving moderate inaccuracy, but without trimming and with low chip area [10][22]. In nanometer CMOS, vertical PNP's have very low current gains (<1), and bandgap temperature sensors based on vertical NPNs have been shown to achieve greater accuracy with simpler circuitry [20]. Thermal diffusivity sensors would also seem well suited for such applications and are actively being researched. With all this activity, it is fair to say that even after more than 25 years of research [46], the field of smart temperature sensors is still quite hot!

References